

KLASIFIKASI KELAYAKAN PEMBERIAN KREDIT NASABAH BANK XYZ MENGGUNAKAN METODE ALGORITMA C4.5 DAN NAIVE BAYES

Yusfina Susanti Ripka Igo¹⁾, Abdul Aziz²⁾, Moh. Ahsan³⁾

^{1,2,3)} Teknik Informatika, Universitas PGRI Kanjuruhan Malang, Jl.S.Supriadi No 48 Bandungrejosari Malang
email : igoyustin10@gmail.com¹⁾, abdul.aziz@unikama.ac.id²⁾, ahsan@unikama.ac.id³⁾

Abstrak

Perbankan memberi layanan kredit untuk memudahkan masyarakat dalam menjalankan usaha. Untuk meminimalkan risiko, bank perlu melakukan analisis kelayakan pemberian kredit kepada nasabah. Penelitian ini bertujuan buat memperoleh tata cara yang terbaik buat klasifikasi kelayakan pemberian kredit kepada nasabah. Alasannya memilih metode terbaik yaitu metode yang layak digunakan dalam klasifikasi pemberian kredit dengan nilai akurasi tertinggi. Akurasi tertinggi dipilih melalui perbandingan algoritma c4.5 dan naive bayes. Hasil akurasi yang didapatkan dari algoritma c4.5 dengan tiga kali pengujian berturut-turut adalah 65,75%, 67,70%, 64,95%, sedangkan naive bayes menghasilkan akurasi 64,72%, 66,67, 63,40%.

Kata Kunci :

Kelayakan pemberian kredit, klasifikasi, algoritma c4.5, naive bayes

Abstract

Banks provide credit services to facilitate the community in running a business. To minimize risk, banks need to analyze the feasibility of providing credit to customers. This study aims to obtain the best method for classifying creditworthiness to customers. The reason for choosing the best method is the method that is feasible to use in the classification of credit granting with the highest accuracy value. The highest accuracy is chosen through a comparison of the c4.5 and naive bayes algorithms. The accuracy results obtained from the c4.5 algorithm with three consecutive tests are 65.75%, 67.70%, 64.95%, while Naive Bayes results in an accuracy of 64.72%, 66.67, 63.40%.

Keywords :

Credit eligibility, classification, c4.5 algorithm, naive bayes

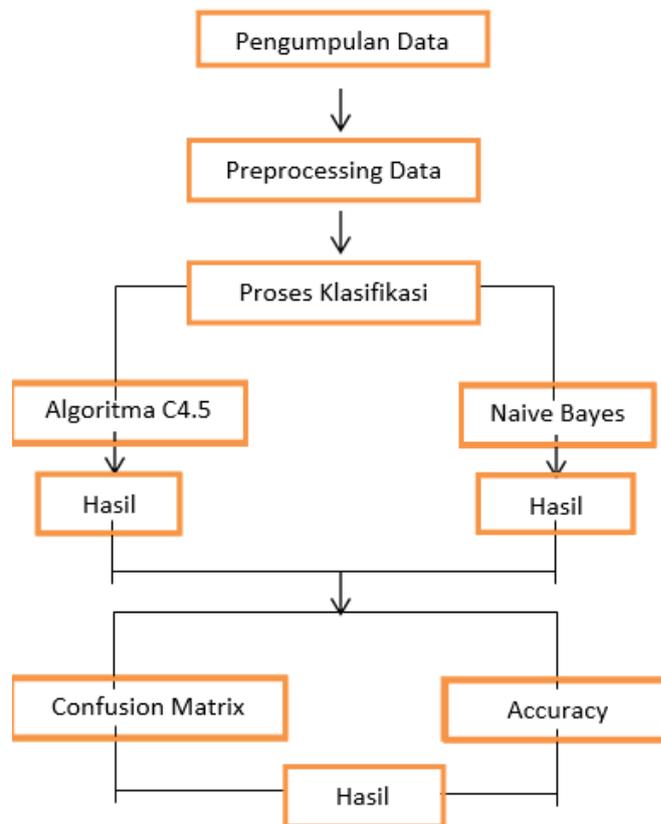
1. PENDAHULUAN

Perbankan memberi layanan kredit agar bisa memudahkan masyarakat dalam menjalankan usaha. Sebelum memberikan kredit harus dilakukan analisis calon nasabah untuk menentukan kelayakan mendapatkan kredit. Hal itu meliputi usia, jumlah angsuran, jumlah pendapatan, jumlah pengeluaran, pekerjaan, jangka waktu, nilai jaminan, jaminan [1].

Kelebihan algoritma C4.5 mampu menangani atribut yang kosong (missing value). Selain itu juga mampu menangani atribut yang kontinu dan mampu memangkas pohon keputusan untuk mengatasi overfitting [2]. Algoritma naive bayes memiliki keunggulan mudah digunakan dan hanya membutuhkan satu kali iterasi data training. Tetapi naive bayes membutuhkan pengetahuan awal untuk mengembalikan keputusan [3].

2. METODE / ALGORITMA

Dalam penelitian penerapan data mining memakai tata cara algoritma C4. 5 serta naive bayes buat mengklasifikasi kelayakan pemberian kredit pada nasabah.



Gambar 2.1 Alur Proses Kelayakan Kredit Nasabah

2.1 Pengumpulan Data

Proses pengumpulan data dalam penelitian ini adalah penelitian meminta data peminjaman di Bank Z baik itu dalam bentuk *hardfile*. Atribut yang digunakan dalam penelitian meliputi usia, jumlah angsuran, jumlah pendapatan, jumlah pengeluaran, pekerjaan, jangka waktu, nilai jaminan, jaminan.

2.2 Preprocessing

Dalam tahapan ini penelitian tidak perlu melakukan preprocessing karena data yang digunakan sudah bersih dan rapih.

2.3 Proses Klasifikasi menggunakan Algoritma C4.5

Pada tahap ini dilakukan perhitungan algoritma C4.5 untuk melakukan pengklasifikasian pada data testing. Untuk mengklasifikasi pengajuan kredit algoritma C4.5 digunakan untuk membentuk pohon keputusan yang mempresentasikan aturan dalam klasifikasi [5].

2.4 Proses Klasifikasi menggunakan Naive Bayes

Pada tahap ini dilakukan perhitungan naive bayes untuk melakukan pengklasifikasian pada data testing. Untuk mengklasifikasi pengajuan kredit terlebih dahulu melakukan perhitungan probabilitas pada data training [6].

2.5 Confusion Matrix

Confusion Matrix adalah sebuah metode yang digunakan untuk menghitung akurasi didalam konsep data mining [7].

Tabel 1. Rumus Confusin Matrix

Confusion Matrix	True	False
Positive	TP (True Positive)	FN (False Negative)
Negative	FP (False Positive)	TN (True Negative)

Keterangan tabel 1 rumus Confusin Matrix

- a. TP (True Positive) adalah jumlah data positif yang diklasifikasi dengan benar
- b. FN (False Negative) adalah jumlah data negative yang diklasifikasi salah
- c. FP (False Positive) adalah jumlah data positive yang diklasifikasi salah
- d. FP (False Positive) adalah jumlah data positive yang diklasifikasi salah
- e. TN (True Negative) adalah jumlah data negative yang diklasifikasi benar

3. HASIL DAN PEMBAHASAN

3.1 Kebutuhan Sistem

Pengumpulan Data

Pengumpulan informasi yang diperoleh dari riset tersebut ialah informasi nasabah Z buat memastikan kelayakan kredit. Jumlah informasi yang diperoleh dari riset tersebut sebanyak 1929 informasi. Kriteria yang digunakan dalam riset ini sebanyak 11 atribut antara lain ialah usia, pemasukan keluarga, pengeluaran keluarga, pinjaman, keahlian angsuran, jangka waktu, nilai jaminan, pekerjaan, anggungan, status anggungan, tujuan serta kolektibilitas.

Preprocessing Data

Pada tahap ini tidak perlu dilakukan preprocessing data karena data yang didapatkan memiliki kelengkapan atribut dan data yang didapatkan rapi.

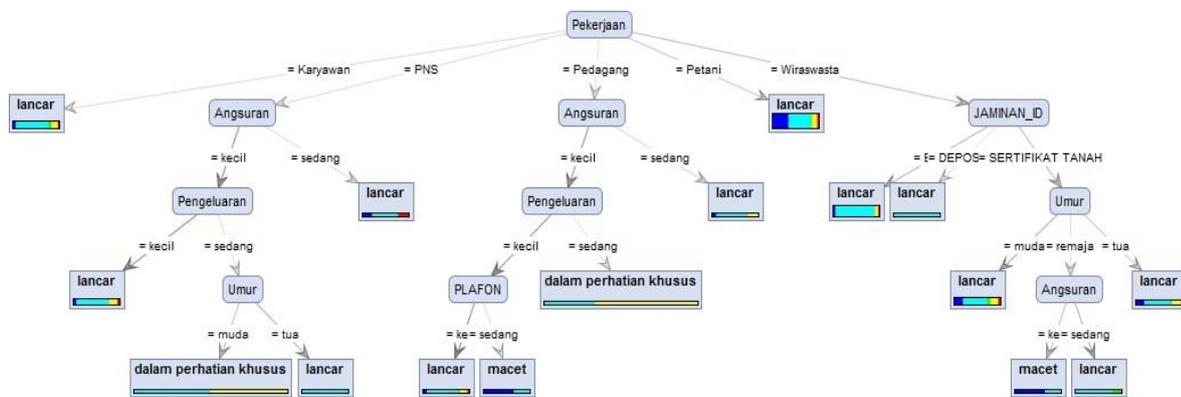
Perhitungan Algoritma C4.5

Hitung nilai entropy dari data training dikenal jumlah permasalahan adalah 1929 yang beresiko lancar 1270, beresiko dalam perhatian khusus 235, beresiko kurang lancar 56, beresiko diragukan 55 dan yang beresiko macet 313. Sehingga didapatkan entropy:

$$\text{Entropy (S)} = \sum n-p_i * \log_2 p_i \tag{1}$$

No	Atribut		Jumlah kasus (S)	Lancar	Dalam perhatian khusus	Kurang lancar	Diragukan	Macet	Entropy	Gain
	Total		1929	1270	235	57	54	313	0,717	0
1	Umur	Remaja	51	0	0	0	0	0	0,026	0,691
		Muda	1237	0	0	0	0	0	0,641	0,076
		Tua	640	0	0	0	0	0	0,029	0,688
2	Pendapatan	Kecil	1487	0	0	0	0	0	0,693	0,024
		Sedang	378	0	0	0	0	0	0,175	0,542
		Besar	64	0	0	0	0	0	0,029	0,688
	Pengeluaran	Kecil	1552	0	0	0	0	0	0,804	-0,087
		Sedang	338	0	0	0	0	0	0,157	0,56

		Besar	39	0	0	0	0	0	0,018	0,699
2										
4	Pinjaman	Kecil	1542	0	0	0	0	0	0,79	-0,073
		Sedang	253	0	0	0	0	0	0,117	0,6
		Besar	134	0	0	0	0	0	0,062	0,655
5	Angsuran	Kecil	1495	0	0	0	0	0	0,697	0,02
		Sedang	374	0	0	0	0	0	0,193	0,524
		Besar	60	0	0	0	0	0	0,027	0,69
6	Nilai jaminan	Kecil	1524	0	0	0	0	0	0,790	-0,073
		Sedang	195	0	0	0	0	0	0,101	0,616
		Besar	210	0	0	0	0	0	0,108	0,609
	Pekerjaan	Wiraswasta	685	0	0	0	0	0	0,319	0,398
		Karyawan	178	0	0	0	0	0	0,092	0,625
		Pedagang	227	0	0	0	0	0	0,105	0,612



Gambar 3.1 Pohon Keputusan Menggunakan Training 80%

Pekerjaan = Karyawan : lancar { macet=14, lancar=116, diragukan=11, dalam perhatian khusus=23, kurang lancar=4 }

- Pekerjaan = PNS

|| Angsuran = kecil

| | | Pengeluaran = kecil: lancar { macet=10, lancar=84, diragukan=1, dalam perhatian khusus=21, kurang lancar=6 }

| | | Pengeluaran = sedang

| Umur = muda: dalam perhatian khusus { macet=0, lancar=1, diragukan=0, dalam perhatian khusus=1, kurang=0 }

| Umur = tua: lancar { macet=0, lancar=3, diragukan=0, dalam perhatian khusus=0, kurang=0 }

Klasifikasi Kelayakan Pemberian Kredit Nasabah

Bank XYZ Menggunakan Metode Algoritma C4.5 Dan Naive Bayes

|| Angsuran = sedang: lancar {macet=1, lancar=6, diragukan=0, dalam perhatian khusus=2, kurang=0}

Pekerjaan = Petani: lancar {macet=193, lancar=259, diragukan=12, dalam perhatian khusus=56, kurang lancar=18}

- Pekerjaan = Wiraswasta

| JAMINAN_ID = BPKB: lancar {macet=22, lancar=275, diragukan=2, dalam perhatian khusus=21, kurang lancar=7}

| JAMINAN_ID = DEPOSITO: lancar {macet=0, lancar=8, diragukan=0, dalam perhatian khusus=0, kurang lancar=0}

| JAMINAN_ID = SERTIFIKAT TANAH

|| Umur = muda: lancar {macet=41, lancar=100, diragukan=15, dalam perhatian khusus=32, kurang lancar=9}

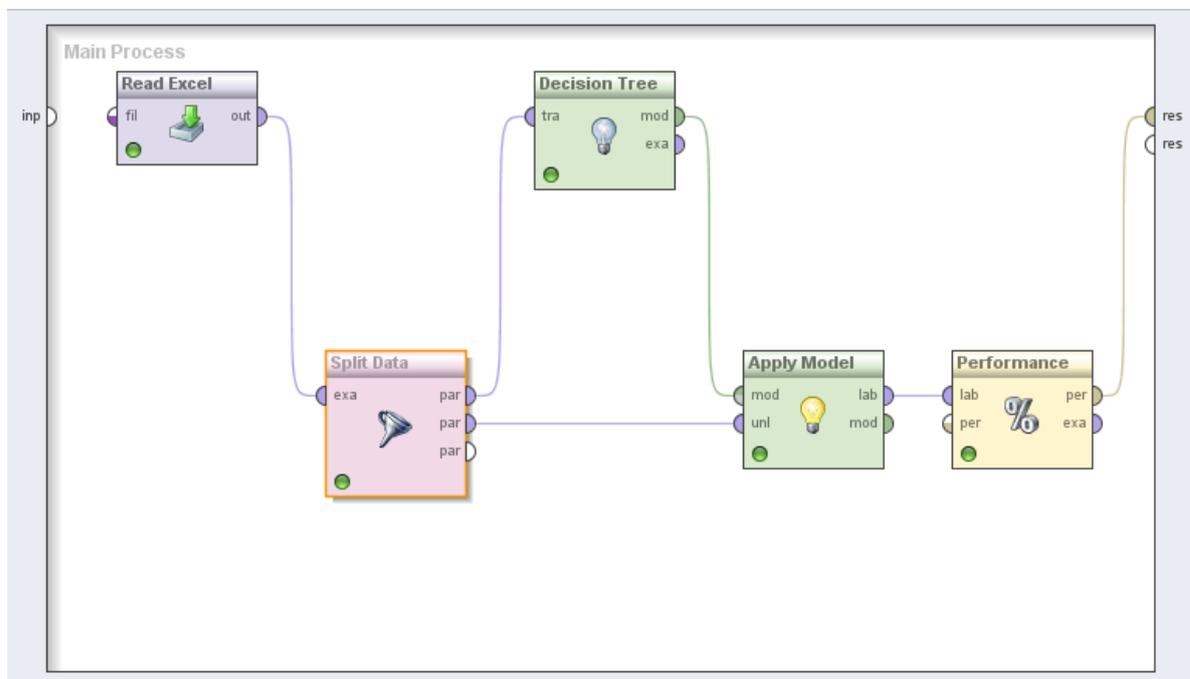
| Umur = remaja

|| Angsuran = kecil: macet {macet=2, lancar=1, diragukan=0, dalam perhatian khusus=0, kurang lancar=0}

|| Angsuran = sedang: lancar {macet=0, lancar=4, diragukan=1, dalam perhatian khusus=0, kurang lancar=0}

| Umur = tua: lancar {macet=19, lancar=52, diragukan=3, dalam perhatian khusus=18, kurang lancar=2}

Gambar 3 merupakan pengujian model algoritma C4.5 menggunakan software Rapid Minne. Read Excel digunakan untuk mengambil data yang sudah kita simpan dalam format excel, Decision Tree yaitu metode yang digunakan, Split data dipakai untuk membagikan data training dan data testing, Apply Model yaitu pembentukan model dan Performance yaitu tool yang digunakan untuk mendapatkan akurasi.



Gambar 3.2 Pengujian Algoritma C4.5

3.2 Evaluasi Confussion Matrix

Tabel 3. Accuracy Hasil Rapiddminer Dengan Data Training 80%

	true macet	true lancar	True diragukan	True dalam perhatian khusus	True kurang lancar	Class predisi on
Pred.macet	0	0	0	0	0	
Pred.lancar	59	262	10	44	12	67,70%
Pred.diragukan	0	0	0	0	0	0,00%
Pred.dalam perhatian khusus	0	0	0	0	0	0,00%
Pred.kurang lancar	0	0	0	0	0	0,00%
Class recall	0,00%	100,00%	0,00%	0,00%	0,00%	

$$Acc = \frac{TM+TL+TD+TDP+TK}{FM+TL+FD+FK+FD}$$

$$Acc = \frac{0+262+0+0+0}{59+262+10+44+12} = \frac{262}{387} = 67,70 \%$$

Sesuai dengan tabel 3 menunjukkan bahwa, tingkat akurasi sesuai dengan data training 1543 dan data testing 386 menggunakan algoritma C4.5 adalah sebesar 67,70%.

3.3 Evaluasi Confusion Matrix

Pengujian dari hasil klasifikasi yang diperoleh dari perhitungan algoritma naive bayes dengan menggunakan metode Confusion Matrix yang terdapat pada tabel dibawah ini.

Tabel 4. Hasil Akurasi Menggunakan Rapidminer

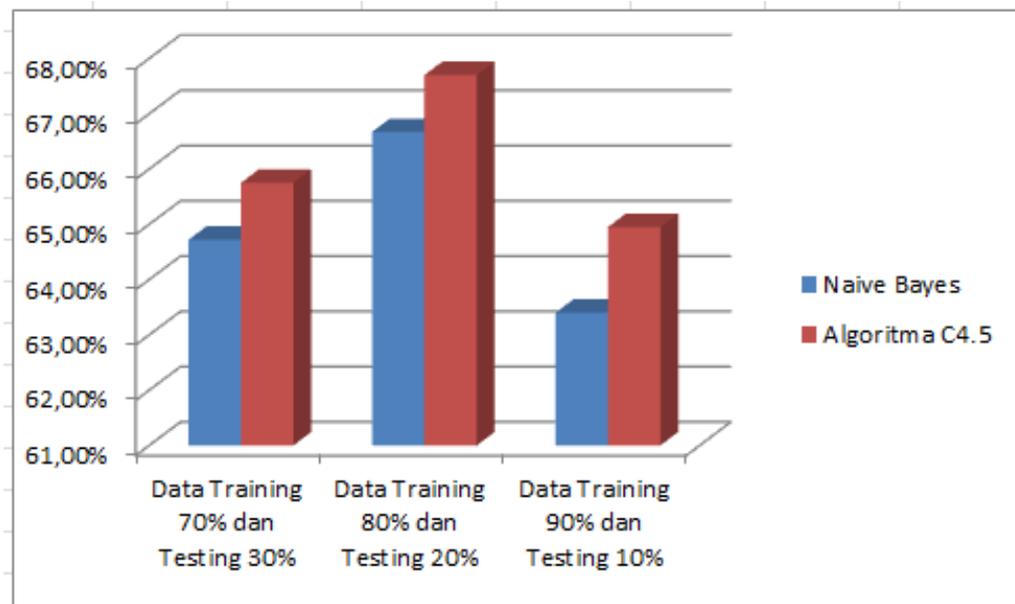
	true macet	true lancar	True diragukan	True dalam perhatian khusus	True kurang lancar	Class predisi on
Pred.macet	3	0	0	0	0	100,00%
Pred.lancar	53	254	9	41	12	68,83%
Pred.diragukan	3	8	1	3	0	6,67%
Pred.dalam perhatian khusus	0	0	0	0	0	0,00%
Pred.kurang lancar	0	0	0	0	0	0,00%
Class recall	5,08%	96,95%	10,00%	0,00%	0,00%	

$$Acc = \frac{TM+TL+TD+TDP+TK}{TM+FM+TL+FD+FD+FK+FM+FL+FD+FD+FK}$$

$$Acc = \frac{3+254+1+0+0}{3+53+254+9+41+12+3+8+1+3} = \frac{258}{387} = 66,66\%$$

Jadi nilai akurasi yang diperoleh dari perhitungan data testing yang diperoleh dari perhitungan naive bayes ini adalah 66,67%. Hasil pengujian akurasi menggunakan metode Confusion Matrix pada rappidminer tingkat akurasi yang diperoleh adalah 66,66% dengan jumlah data yang diklasifikasikan dengan benar pada kategori lancar sebanyak 254 data, dan yang diklasifikasikan dengan benar pada kategori macet sebanyak 3 data. Jumlah data yang diklasifikasikan salah pada kategori lancar sebanyak 0 data dan jumlah data yang diklasifikasi salah pada kategori macet adalah sebanyak 53 data dari 386 data testing yang digunakan pada penelitian.

Terlihat dari gambar 3.3 dapat disimpulkan bahwa hasil perbandingan algoritma C4.5 dan naive bayes yang menghasilkan akurasi yang paling tinggi yaitu algoritma C4.5. Dengan hasil akurasi yang didapatkan sebesar 67,70%.



Gambar 4. Hasil Perbandingan Akurasi Algoritma C4.5 dan Naive Bayes

4. KESIMPULAN

Pengklasifikasian menggunakan metode algoritma C4.5 dengan menggunakan data training dan data testing yang dipilih secara random dengan menggunakan bantuan software Rapidminer. Sedangkan pengklasifikasian menggunakan naive bayes dengan menggunakan data training dan data testing untuk menghitung nilai probabilitas. Hasil perbandingan dari klasifikasi menggunakan kedua metode dapat disimpulkan bahwa metode algoritma C4.5 yang lebih cocok untuk pengklasifikasian data kredit Nasabah Z dengan hasil akurasi sebesar 67,70%. Sedangkan hasil akurasi menggunakan metode naive bayes sebesar 66,67%.

5. REFERENSI

- [1] Anam, C., & Santoso, H. B. (2018). Perbandingan Kinerja Algoritma C4. 5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa. *ENERGY*, 8(1), 13-19.
- [2] Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah*, 10(2), 160-165.
- [3] Annur, H., & Lasulika, M. E. (2019). Klasifikasi Nasabah Kredit Koperasi Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Informatika Upgris*, 5(2).
- [4] Devita, R. N., Herwanto, H. W., & Wibawa, A. P. (2018). Perbandingan kinerja metode naive bayes dan k-nearest neighbor untuk klasifikasi artikel berbahasa indonesia. *JTIK (Jurnal Teknologi Informasi dan Ilmu Komputer)* Vol, 5.
- [5] Fatimah, C. (2017). Hubungan Hukum Antara Bank dan Nasabah Penyimpan Dana Menurut Undang-Undang Perbankan. *Lex Et Societatis*, 5(9).
- [6] Firman Tempola, M. M., & Khairan, A. (2018). Perbandingan Klasifikasi Antara Knn Dan Naive Bayes Pada Penentuan Status Gunung Berapi Dengan K-Fold Cross Validation. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 5(5).
- [7] Fitriani, E. (2020). Perbandingan Algoritma C4. 5 Dan Naive Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan. *SISTEMASI: Jurnal Sistem Informasi*, 9(1), 103-115.